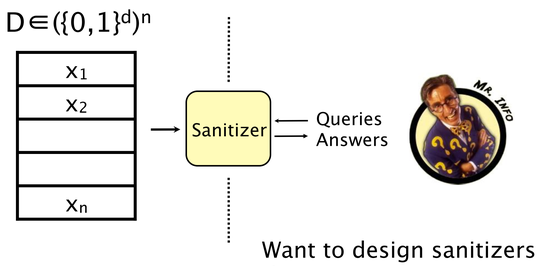
Answering $n^{2+o(1)}$ Counting Queries with Differential Privacy is Hard

Jonathan Ullman
CRCS Postdoctoral Fellow, Harvard University

Privacy Tools for Sharing Research Data
A National Science Foundation Secure and Trustworthy Cyberspace Project

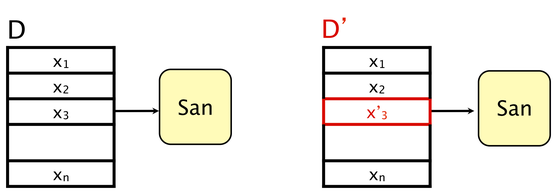
Privacy-Preserving Data Analysis



Want to design sanitizers that are:

- Protect data behind a sanitizer
- differentially private
- accurate
- computationally efficient

Differential Privacy [DMNS]



D and D' are neighbors if they differ on one user's data

San is (ε,δ)-differentially private if for all neighbors D, D' and every S⊆Range(San)

$$\Pr[San(D) \in S] \leq e^{\varepsilon} \cdot \Pr[San(D') \in S] + \delta$$

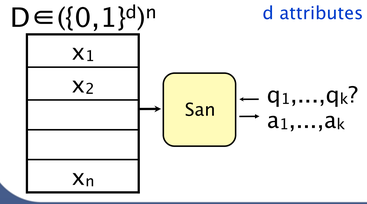
Counting Queries

Counting query: What fraction of records satisfy property q? e.g. q(x) = Att1 OR Att2

D∈({0,1}^d)^n

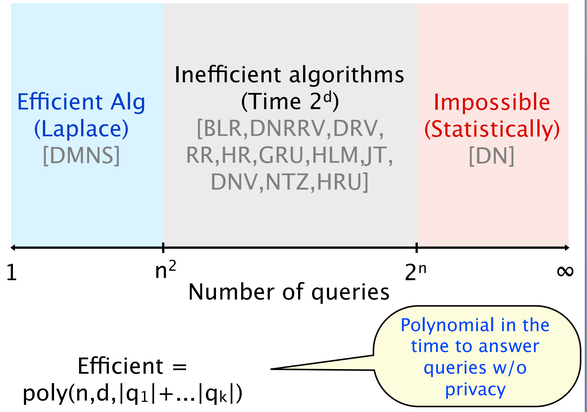
Att 1	Att 2	Att 3	Att 4	
N	N	Y	Y	q(x1)=0
Y	Y	Y	Y	q(x2)=1
N	N	N	N	q(x3)=0
Y	N	N	N	q(x4)=1
				q(D)=1/2

d attributes per record



Accurate if $|q_j(D) - a_j| \leq 1/3$ for every q_j

How Many Queries Can We Answer?



Efficient = poly(n,d,|q1|+...|qk|)

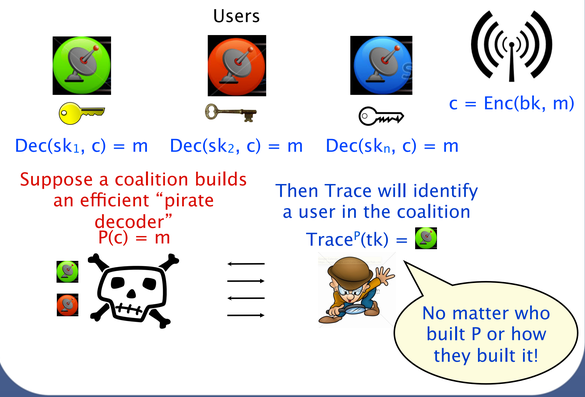
Polynomial in the time to answer queries w/o privacy

Main Result

- There is no efficient, DP sanitizer that accurately answers $n^{2+o(1)}$ arbitrary counting queries. (Assuming OWF)
- ...even if the queries are constant depth circuits. (Assuming more crypto)
- ... proof refines and extends connection between traitor-tracing and DP [DNRRV]
  - Need to construct an unusual new scheme

Traitor-Tracing Schemes [CFN]

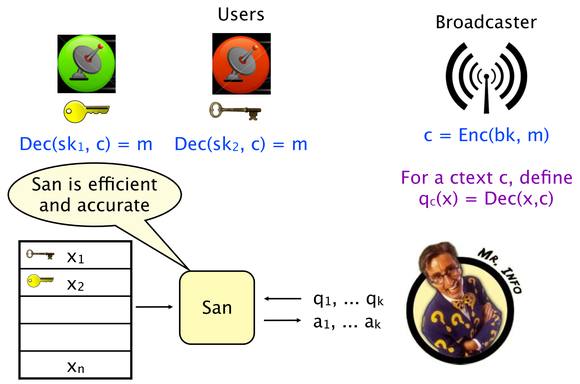
(Gen,Enc,Dec,Trace)



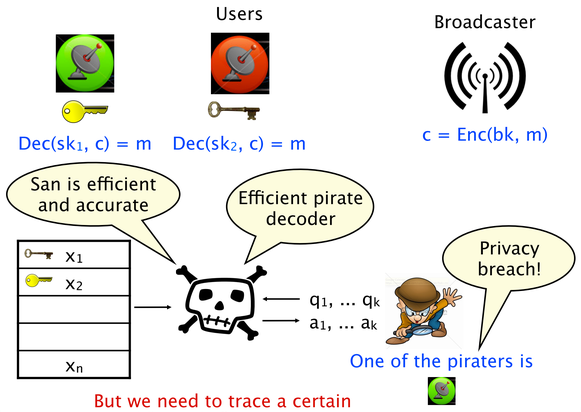
Traitor-Tracing vs. DP [DNRRV]

- Traitor-tracing is "the opposite of privacy"
- Traitor-tracing = Given any algorithm that has the "functionality" of the user keys, the tracer can identify one of its user keys
- Privacy = There is an algorithm that has the "functionality" of the database but no one can identify any of its records

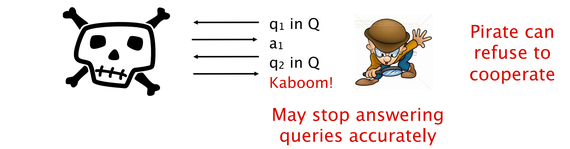
Traitor-Tracing vs. DP [DNRRV]



Traitor-Tracing vs. DP [DNRRV]



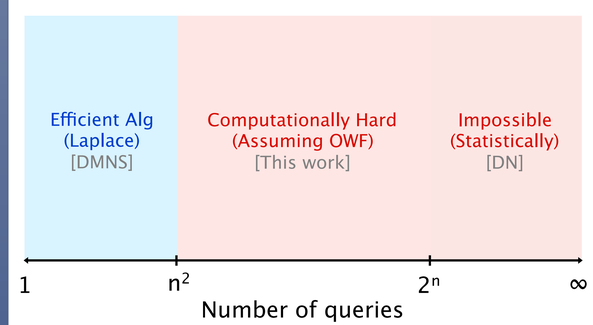
Stateful Pirate Decoders



Tracing Stateful Pirates

- A TTS that traces stateful* pirates with k queries implies no efficient DP alg accurately answers k queries
- Assuming OWF, there is a TTS that can trace stateful* pirates with $n^{2+o(1)}$ queries
  - Construction uses fingerprinting codes [BS] (see Mark's poster)
- ➔ Assuming OWF, there is no efficient DP algorithm that accurately answers $n^{2+o(1)}$ arbitrary counting queries

Conclusion



- Better understanding of traitor-tracing is the key to understanding when efficient differentially private algorithms exist

CRCS Center for Research on Computation and Society at Harvard School of Engineering and Applied Sciences
BERKMAN CENTER FOR INTERNET & SOCIETY AT HARVARD UNIVERSITY
The Institute for Quantitative Social Science at Harvard University
DATA PRIVACY LAB